

Branching into the Unknown: Inferring collective dynamical states from subsampled systems

J. Wilting¹ and V. Priesemann^{1,2}

¹Max-Planck-Institute for Dynamics and Self-Organization, Göttingen

²Bernstein-Center for Computational Neuroscience, Göttingen

August 26, 2016

Abstract

When studying the dynamics of complex systems, one can rarely sample the state of all components. We show that this spatial subsampling typically leads to severe underestimation of the risk of instability in systems with propagation of events. We analytically derived a subsampling-invariant estimator and applied it to non-linear network simulations and case reports of various diseases, recovering a close relation between vaccination rate and spreading behavior. The estimator can be particularly useful in countries with unreliable case reports, and promises early warning if e.g. antibiotic resistant bacteria increase their infectiousness. In neuroscience, subsampling has led to contradictory hypotheses about the collective spiking dynamics: asynchronous-irregular or critical. With the novel estimator, we demonstrated for rat, cat and monkey that collective dynamics lives in a narrow subspace between the two. Functionally, this subspace can combine the different computational properties associated with the two states.

How can we infer properties of a high-dimensional dynamical system if we can only observe a very small part of it? This problem of spatial subsampling is common to almost every area of research where spatially extended, time evolving systems are investigated. For example, in many diseases the number of reported infections may be much lower than the unreported ones[1, 2], or in the financial system only a subset of all banks is evaluated when assessing the risk of developing system wide instability [3] (“stress test”). Spatial subsampling is particularly severe when recording neuronal spiking activity, because the number of neurons that can be recorded with ms precision is vanishingly small compared to the number of all neurons in a brain area[4, 5] (Fig. 1a).

Here, we show for a large class of systems, namely systems that can be approximated by a process with 1st order autoregressive representation (PAR), that subsampling leads to a strong overestimation of their stability. Besides the conventional AR(1) process, PARs include e.g. branching processes [6, 7, 8] and Kesten processes [9]. PARs are ubiquitously used as a first order approximation to assess the self-sustaining dynamics and stability of systems with propagating activity, including epidemic spread of infectious diseases [10, 11], cell proliferation [12], evolution [13] (further biological applications in [14]), neutron processes in nuclear power reactors [15], spread of bankruptcy [16, 17], evolution of stock prices [18, 19], or the propagation of spiking activity in neural networks [20, 21] (Fig. 1b). In these systems, overestimating their stability can fa-

cilitate catastrophes, e.g. when misjudging the risk that a disease becomes epidemic, or that a brain area becomes epileptic. Correct risk prediction is essential to timely initiate counter actions which can mitigate the propagation of events. We analytically derived the origin of the estimation bias and developed a novel estimator, which we analytically proof to be unbiased under subsampling. To assure that a PAR is a reasonable approximation of the evaluated complex system, and to exclude contamination through potential non-stationarities, we included a set of automated, data-driven tests.

In a PAR, the activity in the next time step, A_{t+1} , depends linearly on the current activity A_t . In addition, it allows external input, e.g. drive from stimuli or other brain areas, with a mean rate h , yielding the autoregressive representation

$$\langle A_{t+1} | A_t \rangle = m A_t + h. \quad (1)$$

Here, $\langle \cdot | \cdot \rangle$ denotes the conditional expectation. The stability of A_t is solely governed by m , for example the mean number of spikes triggered by *one* spike in a neuron. The activity is stationary if $m < 1$, while it grows exponentially if $m > 1$. The state $m = 1$ separates the stable from the unstable regime. Especially close to this transition, a correct estimate of m is vital to assess the risk that A_t develops a large, potentially devastating cascade or avalanche of events (e.g. an epileptic seizure or an epidemic disease outbreak), either generically or via a minor increase in m .

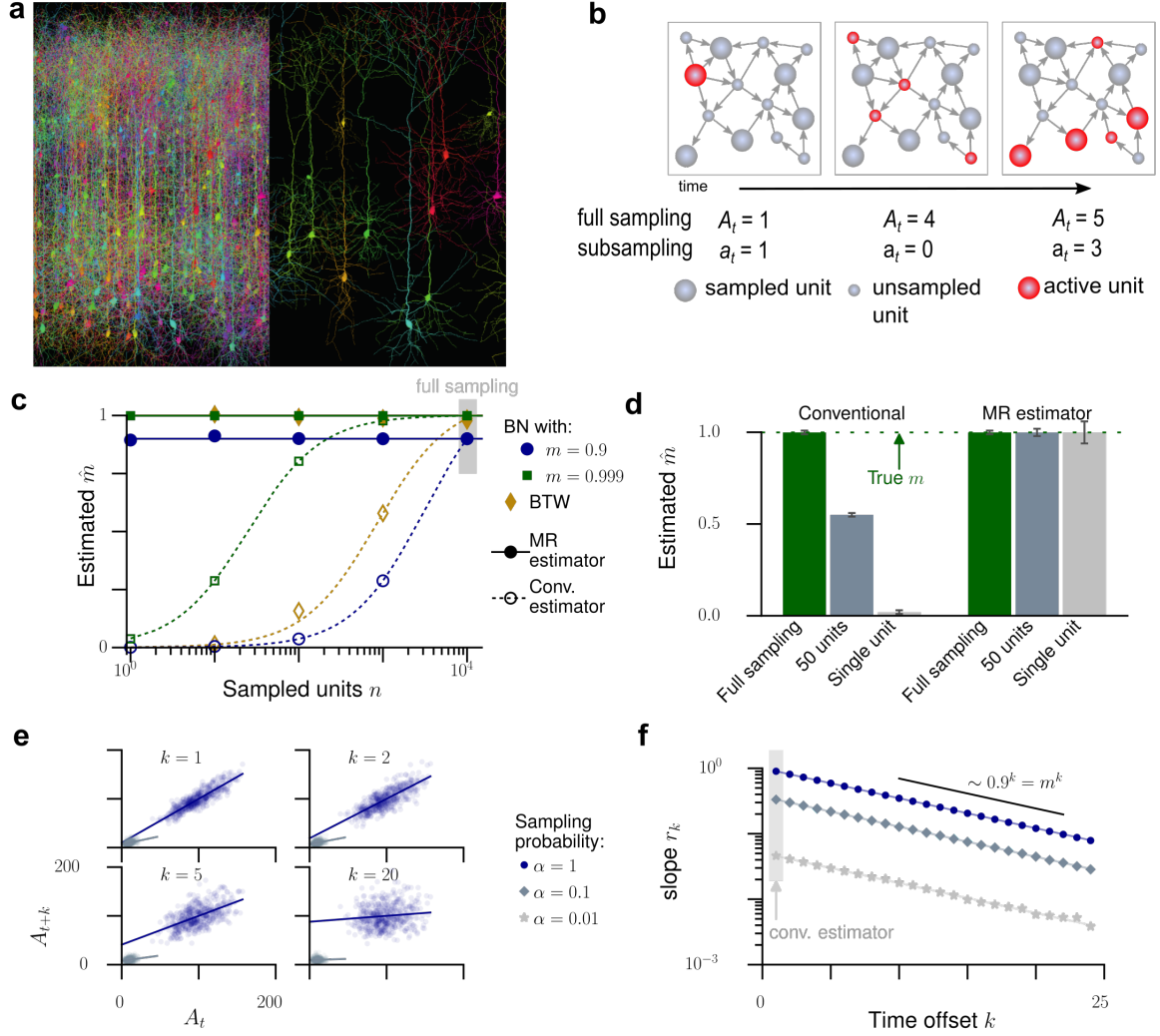


Figure 1: Spatial subsampling. **a.** In complex networks, such as the brain, often only a small subset of all units can be sampled (spatial subsampling); figure created using TREES [22]. **b.** In a branching network (BN), an active unit (e.g. a spiking neuron, infected individual, or defaulting bank) activates some of its neighbors in the next time step. Thereby activity can spread over the system. As the subsampled activity a_t may significantly differ from the actual activity A_t , spatial subsampling can impair inferences about the dynamical properties of the full system. **c.** In recurrent networks (BN, Bak-Tang-Wiesenfeld model / BTW), the conventional estimator (empty symbols) substantially underestimates the branching ratio m when less units n are sampled, as theoretically predicted (dashed lines). The novel MR estimator (full symbols) always returns the correct estimate, even when sampling only 10 or 1 out of all $N = 10^4$ units. **d.** For a BN with $m = 0.999$, the conventional estimator infers $\hat{m} = 0.55$ or $\hat{m} = 0.02$ when sampling 50 or 1 units respectively, in contrast to multistep regression (MR) estimation. **e.** MR estimation is exemplified for a subcritical branching process ($m = 0.9$, $h = 10$), where active units are observed with probability α . Multiple regressions with different time offsets k are shown. Under subsampling (gray), the regression slopes r_k are smaller than under full sampling (blue). **f.** While conventional estimation of m relies on the linear regression r_1 and is biased under subsampling, MR estimation infers \hat{m} from the exponential relation $r_k \propto m^k$, which remains unaffected under subsampling.

Conventional estimation of m uses linear regression [23] (r_1) of activity at time t and $t + 1$, because r_1 directly returns \hat{m} owing to the autoregressive representation. This estimation of m is consistent if the full activity A_t is known [23, 24]. However, under subsampling it can be strongly biased, as we show here. To derive the bias quantitatively, we only assume that the subsampled activity a_t is *in expectation* linear in A_t , $\langle a_t | A_t \rangle = \alpha A_t + \beta$ with two constants α and β (Def. S1). This represents, for example, sampling a fraction α of all neurons in a brain area. Then the conventional estimator is biased by $m(\alpha^2 \text{Var}[A_t] / \text{Var}[a_t] - 1)$ (Corollary S5). The bias vanishes when all units are sampled ($\alpha = 1$, Fig. 1c,d). We emphasize that this bias is inherent to subsampling and cannot be overcome by obtaining longer recordings.

The novel estimator takes a different approach (Def. S2). Instead of directly using the biased regression r_1 , we make use of our analytical result that all regressions r_k between a_t and a_{t+k} are biased *by the same factor* $b = \alpha^2 \text{Var}[A_t] / \text{Var}[a_t]$ (Fig. 1e, Theorem S4). Hence, the exponential relation [25] $r_k = m^k$ (Theorem S1) under full sampling translates to

$$r_k = \alpha^2 \frac{\text{Var}[A_t]}{\text{Var}[a_t]} m^k = b m^k \quad (2)$$

under subsampling. The factor b is, in general, not known. However, one does not need to know it to estimate m from the exponential relation, which serves as the heart of our new multiple regression (MR) estimator (Fig. 1f, Corollary S2 and Theorem S4).

The MR estimator is unbiased under subsampling, because the system itself evolves independently of the sampling process: While subsampling biases each regression r_k by decreasing the mutual dependence between subsequent observations (a_t, a_{t+k}), the temporal decay in $r_k \sim \hat{m}^k = e^{-k/\tau}$ remains unaffected. Here, $\tau = -(\log m)^{-1}$ refers to the autocorrelation time of stationary (subcritical) processes, where autocorrelation and regression r_k are equal. Thus for subcritical PARs, subsampling decreases the autocorrelation *strength* r_k , while the autocorrelation *time* τ is preserved. Making use of this result allows for an unbiased estimate of m even when sampling only a single unit (Fig. 1f).

PARs are typically only a first order approximation of real world event propagation. However, their mathematical structure allowed for an analytical derivation of the subsampling bias and the unbiased estimator. To show that the MR estimator returns correct results also for more complex systems, we applied it to more complex simulated systems: a branching network [21] (BN) and the non-linear Bak-Tang-Wiesenfeld model [26] (BTW). In contrast to generic PARs, these models (a) run on recurrent networks and (b) are of finite size. In addition, the second model shows (c) completely deterministic propagation of activity instead of the stochastic propagation that characterizes PARs, and (d) the activity of each unit depends of many past

time steps, not only one. Both models approximate neural activity propagation in the cortex [20, 21, 4, 5, 27, 28]. In both models, the numerical estimates very well match the analytically derived bias (dashed lines in Fig. 1c, see Eq. S3), although the BN and BTW are only approximated by a PAR. The bias is considerable: For example, sampling 10% or 1% of the neurons in a BN with $m = 0.9$ resulted in the wrong estimates $\hat{m} = r_1 = 0.312$, or even $\hat{m} = 0.047$, respectively. Thus a process fairly close to instability ($m = 0.9$) is mistaken as Poisson-like ($\hat{m} = 0.047 \approx 0$) just because sampling is constrained to 1% of the neurons. Thereby the risk that systems may develop instabilities is tremendously underestimated.

MR estimation is readily applicable to subsampled data, because it only requires a sufficiently long time series a_t , and the assumption that in expectation a_t is proportional to A_t . Hence, in general it suffices to sample the system randomly, without even knowing the system size N , the number of sampled units n , nor any moments of the underlying process. Importantly, one can obtain an unbiased estimate of m , even when sampling only a very small fraction of the process, under homogeneity even when sampling only one single unit (Figs. 1c,d). This robustness makes the estimator readily applicable to any system that can be approximated by a PAR.

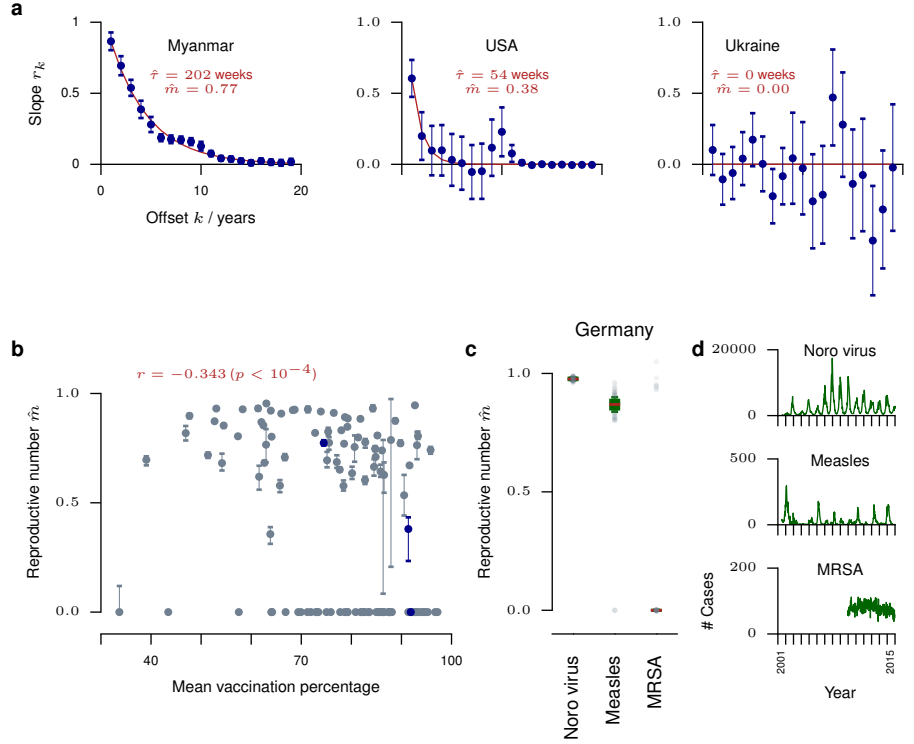
Application to disease case reports. We used the novel estimator to estimate the “reproductive number” m for incidence time series of different diseases. Here, m determines the disease spreading behavior and has been deployed to predict the risk of epidemic outbreaks [11]. However, the problem of subsampling or *under-ascertainment* has always posed a challenge [1, 2]. Our MR estimator promises to infer the reproductive number m directly from the subsampled time series, without even knowing the degree of under-ascertainment.

First, we investigated measles, a contagious infection and a leading cause of death among children, because the vaccination percentage for measles differs in each country, and this is expected to impact the spreading behavior through m . We evaluated worldwide case and vaccination reports for 124 countries provided by the WHO since 1980 (Fig. 2a, see Supp. 8). The reproductive number \hat{m} ranged between 0 and 0.93, and in line with our prediction clearly decreased with increasing vaccination percentage in the respective country (Spearman rank correlation: $r = -0.343, p < 10^{-4}$, Fig. 2b), even though the time series comprised only up to 35 data points per country.

Second, we estimated the reproductive numbers for three under-ascertained [2, 29] diseases in Germany with highly different infectiousness: norovirus, measles, and invasive methicillin-resistant *Staphylococcus aureus* (MRSA, an antibiotic-resistant germ classically associated with health care facilities [30, 31], Figs. 2c,d), and quantified for the first time their propagation behavior. MR estimation returned the highest $\hat{m} = 0.98$ for

Figure 2: Disease propagation.

In epidemic models, the reproductive number m can serve as an indicator for the infectiousness of a disease within a population, and predict the risk of large incidence bursts. We have estimated \hat{m} from incidence time series of measles infections for 124 countries worldwide (see Supp. 7); as well as norovirus, measles, and invasive meticillin-resistant *Staphylococcus aureus* (MRSA) infections in Germany. **a.** MR estimation of \hat{m} is shown for measles infections in three different countries. Errorbars here and in all following figures indicate 1SD or the corresponding 16% to 84% confidence intervals if asymmetric. **b.** The reproductive numbers \hat{m} decrease with the vaccination rate (Spearman rank correlation: $r = -0.342, p < 10^{-4}$). **c.** Reproductive numbers \hat{m} for norovirus, measles and MRSA infections in Germany. **d.** Weekly case report time series for these infections surveyed by the Robert-Koch-Institute.



norovirus, compliant with its high infectiousness [32]. For measles we found an intermediate $\hat{m} = 0.88$, reflecting the vaccination rate of about 97%. For MRSA we identified $m = 0$, confirming that transmission is still minor in Germany [33]. However, a future increase of transmission is feared and would pose a major public health risk [34]. Such an increase could be detected by our estimator, even in countries where case reports are incomplete.

Spiking activity *in vivo* is subcritical. We applied the MR estimator to *in vivo* spiking activity of cortical networks to investigate two contradictory hypothesis about collective spiking dynamics. One hypothesis suggests that the collective dynamics is “asynchronous irregular” (AI), i.e. neurons spike independently of each other and in a Poisson manner ($m = 0$), e.g. reflecting a balanced state [35, 36, 37]. The other hypothesis suggests that neuronal networks operate at criticality ($m = 1$) [20, 38, 39, 40, 41], thus in a particularly sensitive state close to a phase transition. These different hypotheses have distinct implications for the coding strategy of the brain: Criticality is characterized by long-range correlations in space and time, and in models optimizes performance in tasks that profit from long reverberation of the activity in the network [42, 21, 43, 44]. In contrast, the typical balanced state minimizes redundancy [45, 46, 47] and supports fast network responses [35].

Analyzing *in vivo* spiking activity from Macaque monkey prefrontal cortex during a memory task [48], anesthetized cat visual cortex with no stimulus [49], and rat hippocampus during a foraging task [50, 51] returned \hat{m} to be between 0.963 and 0.998 (median $\hat{m} = 0.984$), corresponding to auto-correlation times τ between 100 ms and 2000 ms (Fig. 3d, Fig. S3). This clearly suggests that spiking activity *in vivo* is neither AI-like ($m = 0$), nor consistent with a critical state ($m = 1$), which shows $\tau \rightarrow \infty$. Instead cortical activity lives in a narrow subspace covering only 5% of the physiologically plausible range ($m \in [0, 1]$, as instability is not physiologically sustainable).

The MR estimator can distinguish these states with the necessary precision, because tiny changes in m have large impact on the network dynamics if the network is close to $m = 1$. In fact, the estimator would allow for 100 times higher precision when distinguishing critical from non-critical BN, assuming *in vivo*-like subsampling (sampling $n = 100$ from $N = 10^4$ neurons, Fig. 3d). With larger N , this discrimination becomes even more sensitive (detailed error estimates: Fig. S2 and Supp. 6). As the number of neurons in a given brain area is typically much higher than $N = 10^4$ in the simulation, finite size effects are not likely to account for the observed deviation from criticality $\epsilon = 1 - m \approx 10^{-2}$ *in vivo*, supporting that in rat, cat, and monkey the brain does not operate in a critical state.

The *in vivo* recordings showed clear fluctuations for

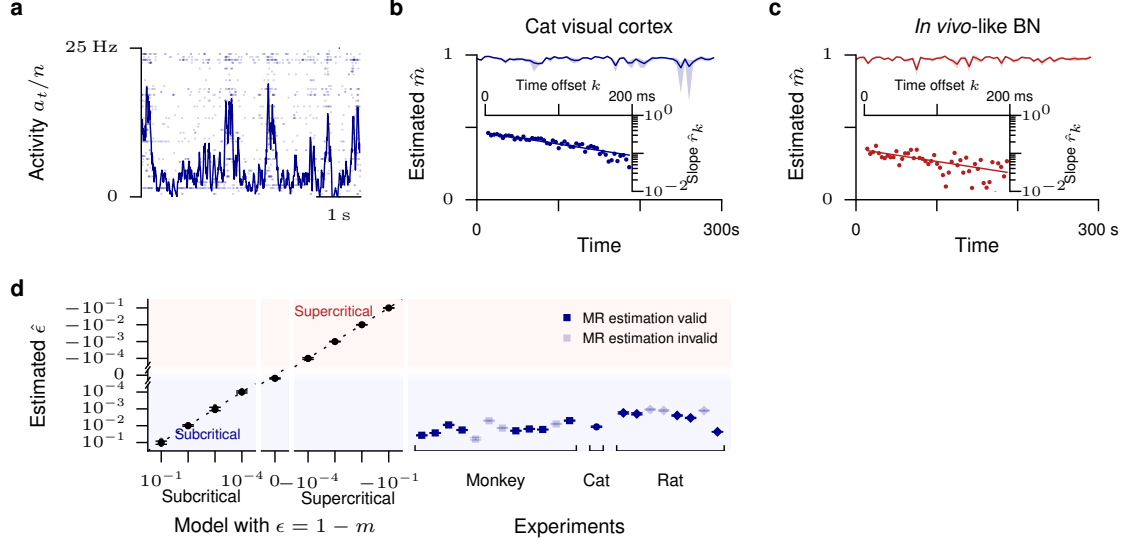


Figure 3: Animal spiking activity *in vivo*. In neuroscience, m denotes the mean number of spikes triggered by one spike. We estimated \hat{m} from spiking activity recorded *in vivo* in monkey prefrontal cortex, cat visual cortex, and rat hippocampus. **a.** Raster spike plot and population rate a_t illustrated for cat visual cortex. **b.** The estimated branching parameter \hat{m} for 59 windows of 5 s length suggests stationarity of m over the entire recording (shaded area: 16% to 84% confidence intervals). Inset: MR estimation for one example window. **c.** A matched, *in-vivo*-like BN simulation showed comparable variation in \hat{m} over consecutive windows. **d.** MR estimation returned the correct distance to instability (criticality) $\epsilon = 1 - m$ for all simulations (subcritical BN to supercritical BP, see Supp. 7). All *in vivo* spike recordings from rat, cat, and monkey, returned clearly a subcritical branching parameter $\hat{m} = 1 - \hat{\epsilon}$, with $0.963 < \hat{m} < 0.998$, (median $\hat{m} = 0.984$, errorbars indicate 16% to 84% confidence intervals). Opaque symbols indicate that MR estimation was rejected (Fig. S3, see Supp. 5).

the population spiking activity (Fig. 3a). Such fluctuations could in principle arise from non-stationarities, which could in turn lead to misestimation of m . This is unlikely for three reasons: First, our MR estimator rejects recordings that show any signature of common non-stationarities (Fig. S1, see Supp. 5). Second, recordings in cat visual cortex were acquired in absence of any stimulation, excluding stimulus related non-stationarities. Third, when splitting the spike recording into short windows, the window-to-window variation of \hat{m} in the recording did not differ from that of a stationary *in vivo*-like BN ($p = 0.3$, Figs. 3b,c). The *in vivo*-like BN was set up with the same branching ratio m , spike rate $\langle a_t \rangle$, number of sampled neurons n , and duration as the experimental recording (for the cat $n = 50$, $m = 0.98$, $\langle a_t \rangle/n = 7.9$ Hz, recording of 295 s length).

We want to stress, that cortical dynamics is more complicated than a simple PAR, because heterogeneity of neuronal morphology and function, non-trivial network topology, and the complexity of neurons themselves are likely to have a profound impact onto the population dynamics. Nevertheless, we found an exponential relation $r_k = b m^k$ for the population activity of all considered species (insets of Fig. 3b,c, Fig. S3), indicating that cortical dynamics can at least to a large extent be approximated by a PAR, despite the diversity of the underlying system. Because of the analytical tractability of such generic models, this analysis allows valuable insight into reverberant dynamics and

the stability of cortical activity. Future analysis can then study the effect of relevant, additional parameters. In order to test for the applicability of a PAR approximation, we defined several tests (see Supp. 5) and included only those experiments, where the approximation by a PAR was considered appropriate (14 out of 21, Fig. S3). Note, that these tests were defined very conservative. For example, we excluded all experiments that showed an offset in the slopes r_k , because this offset is, strictly speaking, not explained by a PAR and might indicate non-stationarities.

Using the novel MR estimator, we have shown that *in vivo* spiking dynamics in three different species operates neither at criticality, nor in an AI state, but in a regime with finite correlation times of hundreds of milliseconds, in agreement with the results by [52]. In summary, because of its broad applicability, in particular because it requires only the knowledge of the recorded time series, but no assumption about the size of the full system, the number of sampled units, or any of the moments of the full process, we expect that MR estimation can substantially contribute to the understanding of real-world dynamical systems in diverse fields of research where subsampling prevails.

Supplementary material

Supp. 1 Applicability of MR estimation

We here analytically derive the novel MR estimator for branching processes (BP) [6, 7, 8]. Analogous derivations apply to any process with a first order autoregressive representation (PAR) [53], because these processes fulfill Eq. (S4). Beside BPs, PARs include autoregressive AR(1) processes, integer-valued autoregressive INAR(1) processes [54] rounded integer-valued autoregressive RINAR(1) processes [55], and Kesten processes [9].

We emphasize that the MR estimator only requires the subsampled recording a_t of a system with full activity A_t conforming with the definition below. It is not necessary to know either the full system size, the number of subsampled units, nor any of the moments of the full process A_t .

Supp. 2 Branching processes

In a branching process (BP) with immigration [6, 7, 8] each unit i produces a random number $y_{t,i}$ of units in the subsequent time step. Additionally, in each time step a random number h_t of units immigrates into the system (drive). Mathematically, BPs are defined as follows [6, 7]: Let $y_{t,i}$ be independently and identically distributed non-negative integer-valued random variables following a law Y with mean $m = \langle Y \rangle$ and variance $\sigma^2 = \text{Var}[Y]$. Further, Y shall be non-trivial, meaning it satisfies $P[Y = 0] > 0$ and $P[Y = 0] + P[Y = 1] < 1$. Likewise, let h_t be independently and identically distributed non-negative integer-valued random variables following a law H with mean rate $h = \langle H \rangle$ and variance $\xi^2 = \text{Var}[H]$. Then the evolution of the BP A_t is given recursively by

$$A_{t+1} = \sum_{i=1}^{A_t} y_{t,i} + h_t, \quad (\text{S1})$$

i.e. the number of units in the next generation is given by the offspring of all present units and those that were introduced to the system from outside.

The stability of BPs is solely governed by the mean offspring m . In the subcritical state, $m < 1$, the population converges to a stationary distribution A_∞ with mean $\langle A_\infty \rangle = h/(1-m)$. At criticality ($m = 1$), A_t asymptotically exhibits linear growth, while in the supercritical state ($m > 1$) it grows exponentially.

Supp. 3 Subsampling

To derive the MR estimator for subsampled data, subsampling is implemented in a parsimonious way, according to the following definition:

Definition S1 (Subsampling). Let $\{A_t\}_{t \in \mathbb{N}}$ be a BP and $\{a_t\}_{t \in \mathbb{N}}$ a sequence of random variables. Then $\{a_t\}_{t \in \mathbb{N}}$ is called a subsampling of $\{A_t\}_{t \in \mathbb{N}}$ if it fulfills the following three conditions:

- (i) Let $t', t \in \mathbb{N}$, $t' \neq t$. Then the conditional random variables $(a_t | A_t = j)$ and $(a_{t'} | A_{t'} = l)$ are independent for any $j, l \in \mathbb{N}$. If $A_t = A_{t'}$ then $(a_t | A_t = j)$ and $(a_{t'} | A_{t'} = j)$ are identically distributed.
- (ii) Let $t \in \mathbb{N}$. Conditioning on a_t does not add further information to the process: The two random variables $(A_{t+1} | A_t = j, a_t = l)$ and $(A_{t+1} | A_t = j)$ are identically distributed for any $j, l \in \mathbb{N}$.
- (iii) There are constants $\alpha, \beta \in \mathbb{R}$, $\alpha \neq 0$, such that $\langle a_t | A_t = j \rangle = \alpha j + \beta$ for all $t, j \in \mathbb{N}$.

Thus the subsample a_t is constructed from the full process A_t based on the three assumptions: (i) The sampling process does not interfere with itself, and does not change over time. Hence the realization of a subsample at one time does not influence the realization of a subsample at another time, and the conditional *distribution* of $(a_t | A_t)$ is the same as $(a_{t'} | A_{t'})$ if $A_t = A_{t'}$. However, even if $A_t = A_{t'}$, the subsampled a_t and $a(t')$ do not necessarily take the same value. (ii) The subsampling does not interfere with the evolution of A_t , i.e. the process evolves independent of the sampling. (iii) *On average* a_t is proportional to A_t up to a constant term.

It will be shown later, that the novel estimator is applicable to any time series a_t that was acquired from a BP conforming with this definition of subsampling. We will demonstrate possible applications at the hand of two examples:

1. Diagnosing infections with probability α . For example, when a BP A_t represents the spread of infections within a population, each infection may be diagnosed with probability $\alpha \leq 1$, depending on the sensitivity of the test and the likelihood that an infected person consults a doctor. If each of the A_t infections is diagnosed independent of the others, then the number of diagnosed cases a_t follows a binomial distribution $a_t \sim \text{Bin}(A_t, \alpha)$. Then $\langle a_t | A_t = j \rangle = \alpha j$ is given by the expected value of the binomial distribution. This implementation of subsampling conforms with the definition, with the sampling probability α and the constant in (iii) being identical here.

2. Sampling a subset of system components. In a different application, a high-dimensional system of interacting units forms the substrate on which activation propagates. Often, the states of a subset units are observed continuously, for example by placing electrodes that record the activity of the same neurons over the entire recording (Fig. 1b). This implementation of subsampling in finite size systems is mathematically approximated as follows: If n out of all N model units are sampled, the probability to sample a_t active units out of the actual A_t active units follows a hypergeometric distribution, $a_t \sim \text{Hyp}(N, n, A_t)$. As $\langle a_t | A_t = j \rangle = j n / N$, this representation satisfies Def. S1 with $\alpha = n / N$. Choosing this special implementation of subsampling allows to derive predictions for the Fano factor under subsampling and the spike count cross correlation. First, evaluate $\text{Var}[a_t]$ further in terms of A_t :

$$\begin{aligned} \text{Var}[a_t] &= \langle \text{Var}[a_t | A_t] \rangle + \text{Var}[\langle a_t | A_t \rangle] \\ &= n \left\langle \frac{A_t}{N} \frac{N - A_t}{N} \frac{N - n}{N - 1} \right\rangle + \text{Var}\left[\frac{n}{N} A_t\right] \\ &= \frac{1}{N} \frac{n}{N} \frac{N - n}{N - 1} (N \langle A_t \rangle - \langle A_t^2 \rangle) + \frac{n^2}{N^2} \text{Var}[A_t] \\ &= \frac{n}{N^2} \frac{N - n}{N - 1} (N \langle A_t \rangle - \langle A_t^2 \rangle) + \left(\frac{n^2}{N^2} - \frac{n}{N^2} \frac{N - n}{N - 1} \right) \text{Var}[A_t]. \end{aligned} \quad (\text{S2})$$

This expression precisely determines the variance $\text{Var}[a_t]$ under subsampling from the properties $\langle A_t \rangle$ and $\text{Var}[A_t]$ of the full process, and from the parameters of subsampling n and N . Using Eq. (S2), we could predict the linear regression slopes \hat{r}_k under subsampling (see Theorem S4, Eq. (S15)) in more detail:

$$r_k = \alpha^2 \frac{\text{Var}[A_t]}{\text{Var}[a_t]} m^k = \frac{n(N - 1) \text{Var}[A_t]}{(N - n)(N \langle A_t \rangle - \langle A_t^2 \rangle) + (nN - N) \text{Var}[A_t]} m^k =: b(N, n, \langle A_t \rangle, \text{Var}[A_t]) m^k. \quad (\text{S3})$$

The term $b = b(N, n, \langle A_t \rangle, \text{Var}[A_t])$ is constant when subsampling a given (stationary) system, and quantifies the factor by which \hat{m} is biased when using the conventional estimate for m . It depends on N , n and the first two moments of A_t and is thus known for a BP. This relation was used for Fig. 1c.

Supp. 4 MR estimation

We here derive an estimator for the mean offspring m based on the autoregressive representation of the BP,

$$\langle A_{t+1} | A_t = j \rangle = m j + h. \quad (\text{S4})$$

This novel estimator is based on multistep regressions [25] (MR estimator), which generalize (S4) to arbitrary time steps k . From iteration of Eq. (S4), it is easy to see that

$$\langle A_{t+k} | A_t = j \rangle = m^k j + h \frac{1 - m^k}{1 - m}. \quad (\text{S5})$$

Definition S2 (Multistep regression estimator). Consider a subsampled BP $\{a_t\}$ of length T . Let $k_{\max} \in \mathbb{N}$, $k_{\max} \geq 2$. Then multistep regression (of k_{\max} -th order) estimates m in the following way:

1. For $k = 1, \dots, k_{\max}$, estimate the slope \hat{r}_k of linear regression between the pairs $\{(a_t, a_{t+k})\}_{t=0}^{T-k}$, e.g. by least square estimation (Fig. 1e).
2. Based on the relation $r_k = b \cdot m^k$, estimate \hat{b} and \hat{m} by minimizing the sum of residuals

$$R(\hat{b}, \hat{m}) = \sum_{k=1}^{k_{\max}} \left(\hat{r}_k - \hat{b} \cdot \hat{m}^k \right)^2, \quad (\text{S6})$$

with the collection of slopes $\{\hat{r}_k\}_{k=1}^{k_{\max}}$ obtained from step 1 (Fig. 1f).

Then \hat{m} is the multistep regression (MR) estimate of the mean offspring m . For the application to experimental data, we further applied tests to identify nonstationarities (see Supp. 5).

We first prove that the MR estimator is consistent in the fully sampled case, and will then show the consistency under subsampling. First, we need the following result about the individual linear regression slopes \hat{r}_k under full sampling:

Theorem S1. *The slope \hat{r}_k , obtained from A_t under full sampling, is a consistent estimator for m^k . If the process is subcritical, then \hat{s}_k is also a consistent estimator for $h \frac{1-m^k}{1-m}$.*

Remark. For $k = 1$, these results were already obtained by [23, 56, 24], and details can be found in these sources. Based on their proofs, we here show the generalization to k timesteps.

Proof. Let $k \in \mathbb{N}$, $i \in \{0, \dots, k-1\}$. Construct a new random process by starting at time i and taking every k -th time step of the original process A_t . This new process is given by $A_{t'}^{(k,i)} = A_{i+k \cdot t'}$ with the index $t' \in \mathbb{N}$. Hence, the “time” t' of this new process relates to the time t of the old process as $t = i + k \cdot t'$. For a time series of length T , let $r^{(k,i)}$ be the least square estimator for the slope and $\hat{s}^{(k,i)}$ the least square estimator for the intercept of linear regression on all pairs $(A_{t'+1}^{(k,i)}, A_{t'}^{(k,i)})$ from the time series $\{A_{t'}^{(k,i)}\}_{t'=0}^{\lfloor (T-1)/k \rfloor}$. We will derive that $r^{(k,i)}$ is a consistent and unbiased estimator for m^k . According to [24], it is sufficient to show that the evolution of $A_{t'}^{(k,i)}$ can be rewritten as

$$A_{t'}^{(k,i)} = m^k \cdot A_{t'-1}^{(k,i)} + h \frac{1-m^k}{1-m} + \epsilon_{t'}^{(k,i)} \quad (\text{S7})$$

with a martingale difference sequence $\epsilon_{t'}^{(k,i)}$, as this is a stochastic regression equation. Hence, consider

$$\epsilon_{t'}^{(k,i)} = A_{t'}^{(k,i)} - m^k \cdot A_{t'-1}^{(k,i)} - h \frac{1-m^k}{1-m} = A_{i+kt'} - m^k \cdot A_{i+k(t'-1)} - h \frac{1-m^k}{1-m}. \quad (\text{S8})$$

We now show that $(\epsilon_{t'}^{(k,i)})_{t' \in \mathbb{N}}$ is a martingale difference sequence for all k . From iteration of Eq. (S5), it is easy to see that

$$\langle A_{t'}^{(k,i)} | A_{t'-1}^{(k,i)} = j \rangle = \langle A_{kt'+i} | A_{kt'-k+i} = j \rangle = m^k j + h \frac{1-m^k}{1-m} \quad (\text{S9})$$

holds. Hence, $\langle \epsilon_{t'}^{(k,i)} | A_{t'-1}^{(k,i)} = j \rangle = 0$ for any j and $\{\epsilon_{t'}^{(k,i)}\}$ is indeed a martingale difference sequence. Therefore, $\{A_{t'}^{(k,i)}\}_{t'=0}^{\lfloor T/k \rfloor}$ satisfies a linear stochastic regression equation with slope m^k and intercept $h \frac{1-m^k}{1-m}$. The least square estimators return unbiased and consistent estimates for the slope and intercept in the subcritical case, i.e. the estimators converge in probability [23, 56, 24]:

$$\hat{r}^{(k,i)} \xrightarrow{P} m^k \quad \hat{s}^{(k,i)} \xrightarrow{P} h \frac{1-m^k}{1-m}.$$

In the critical and supercritical cases, only $\hat{r}^{(k,i)} \xrightarrow{P} m^k$ holds following [24]. Hence, we obtain $\hat{r}_k \xrightarrow{P} m^k$ for all m and $\hat{s}_k \xrightarrow{P} h(1-m^k)/(1-m)$ if $m < 1$. \square

Corollary S2. *As least square estimation of \hat{b} and \hat{m} from minimizing the residual (S6) is unbiased and consistent, multistep regression is an unbiased and consistent estimator for m under full sampling, $\hat{m} \xrightarrow{P} m$.*

These results were obtained for BPs. However, the derivation is here only based on the autoregressive representation (S4), motivation the following proposition:

Conjecture S3. *Multistep regression is a consistent and unbiased estimator for m for any PAR satisfying Eq. (S4).*

Numerical results for Kesten processes support this conjecture [25].

Next, we show that MR estimation is unbiased in the subcritical case even if only the subsampled a_t is known:

Theorem S4. *Let A_t be a PAR with $m < 1$ and a stationary limiting distribution A_∞ and let the PAR be started in the stationary distribution, i.e. $A_0 \sim A_\infty$. Let a_t be a subsampling of A_t . Multistep regression (MR) on the subsampled a_t is an unbiased estimator of the mean offspring m .*

Proof. The existence of a stationary distribution A_∞ was shown by [7]. The least square estimator for the slope of linear regression is also given by [57]

$$\hat{r}_k = \hat{\rho}_{a_t a_{t+k}} \frac{\hat{\sigma}_{a_t}}{\hat{\sigma}_{a_{t+k}}} \quad (\text{S10})$$

with the the estimated standard deviations $\hat{\sigma}_{a_t}$ and $\hat{\sigma}_{a_{t+k}}$ of a_t and a_{t+k} respectively. In the subcritical state, $\sigma_{a_t} = \sigma_{a_{t+k}}$ because of stationarity. Thus estimating the linear regression slope is equivalent to estimating the Pearson correlation coefficient $\hat{\rho}_{a_t a_{t+k}} = \hat{\rho}_{a_t}(k)$ (which is identical to the autocorrelation function of a_t). In the

following, we calculate the Pearson correlation coefficient for the subsampled time series by evaluating $\langle a_t a_{t+k} \rangle$. We use the law of total expectation in order to express $\langle a_t a_{t+k} \rangle$ not in dependence of a_t , but in terms of A_t :

$$\langle a_t a_{t+k} \rangle = \langle \langle a_t a_{t+k} | A_t, A_{t+k} \rangle \rangle_{A_{t+k}, A_t}, \quad (\text{S11})$$

where the inner expectation value is taken with respect to the joint distribution of a_{t+k} and a_t , and the outer with respect to the joint distribution of A_{t+k} and A_t . Through conditioning on both A_t and A_{t+k} , $(a_t | A_t)$ and $(a_{t+k} | A_{t+k})$ become independent due to Def. S1. Hence, the joint distribution of $(a_t, a_{t+k} | A_t, A_{t+k})$ factorizes, and the expectation value factorizes as well. By definition, $\langle a_t | A_t = j \rangle = \alpha j + \beta$ and hence

$$\langle a_t a_{t+k} \rangle = \langle (\alpha A_{t+k} + \beta) (\alpha A_t + \beta) \rangle_{A_{t+k}, A_t} \quad (\text{S12})$$

Without loss of generality, we here show the proof for $\beta = 0$ which is easily extended to the general case. We express $\langle a_t a_{t+k} \rangle$ in terms of Eq. (S5) using the law of total expectation again:

$$\begin{aligned} \langle a_t a_{t+k} \rangle &= \alpha^2 \langle A_t A_{t+k} \rangle \\ &= \alpha^2 \langle \langle A_t A_{t+k} | A_t \rangle \rangle_{A_t} \\ &= \alpha^2 \langle A_t \left(m^k A_t + h \frac{1-m^k}{1-m} \right) \rangle_{A_t} \\ &= \alpha^2 (m^k \langle A_t^2 \rangle + (1-m^k) \langle A_t \rangle), \end{aligned}$$

where the first expectation was taken with respect to the joint distribution of A_t and A_{t+k} . We then used that $\langle A_t^2 \rangle$ and $\langle A_t \rangle = h/(1-m)$ exist, which follows from stationarity of the process. By a similar argument,

$$\langle a_t \rangle = \langle \langle a_t | A_t \rangle \rangle_{A_t} = \alpha \langle A_t \rangle = \alpha \frac{h}{1-m} \quad (\text{S13})$$

and combining these results the covariance is

$$\text{Cov}[a_{t+k}, a_t] = \langle a_{t+k} a_t \rangle - \langle a_{t+k} \rangle \langle a_t \rangle = \alpha^2 (m^k \langle A_t^2 \rangle + (1-m^k) \langle A_t \rangle^2) - \alpha^2 \langle A_t \rangle^2 = \alpha^2 m^k \text{Var}[A_t]. \quad (\text{S14})$$

Therefore, we find that the estimator \hat{r}_k converges in probability:

$$\hat{r}_k \xrightarrow{P} \rho_{a_t a_{t+k}} = \frac{\text{Cov}[a_{t+k}, a_t]}{\text{Var}[a_t]} = \alpha^2 \frac{\text{Var}[A_t]}{\text{Var}[a_t]} m^k. \quad (\text{S15})$$

Hence, the bias of the conventional estimator \hat{r}_1 is precisely given by the factor $b = \alpha^2 \text{Var}[A_t] / \text{Var}[a_t]$. However, importantly the relation $\hat{r}_k = \hat{b} \hat{m}^k$ still holds for the subsampled a_t . Given a collection of multiple linear regressions $\hat{r}_1, \dots, \hat{r}_{k_{\max}}$, the least square estimation of \hat{b} and \hat{m} from minimizing the residual (S6) yields an unbiased and consistent estimator \hat{m} for the mean offspring m even under subsampling and only requires the knowledge of a_t . \square

This proof also showed that the conventional estimator [23] is biased under subsampling:

Corollary S5. *Let $\{a_t\}$ be a subsampling of a subcritical PAR $\{A_t\}$. Then the classical linear regression estimator \hat{m}_{LR} by [23] is biased by $m \cdot \left(\alpha^2 \frac{\text{Var}[A_t]}{\text{Var}[a_t]} - 1 \right)$.*

Nonstationarity, criticality and supercriticality. Even if the subcritical process is not started in the stationary distribution ($A_0 \sim A_\infty$), the results by [7] show that it will converge to this stationary distribution as $t \rightarrow \infty$. The consistence of the estimator in the fully sampled case is included in our proof of Lemma S1 and follows from the results by [23, 24]. Our proof for the subsampled case (Theorem S4), in contrast, strictly requires stationarity ($A_t \sim A_\infty$ for any t) and the existence of the first two moments of A_t . Nevertheless, numerical results suggest that the MR estimator is also consistent if the subcritical process is not started in the stationary distribution, $A_0 \sim A_\infty$, and also for critical and supercritical cases where no stationary distribution exists (Fig. 3d).

Supp. 5 Identifying Common non-stationarities and Poisson activity.

In many types of analyses, non-stationarities in the time series can lead to wrong results, typically an overestimation of \hat{m} . We developed tests to exclude data sets with signatures of common non-stationarities. The different non-stationarities, their impact on the r_k and the rules for rejection of time series are outlined below.

First, *transient* increases of the drive h_t , e.g. in response to a stimulus, lead to a transient increase in $\langle A_t \rangle$. These transients induce correlations or anti-correlations, which prevail on long time scales (Fig. S1c,d). The autocorrelation function is therefore better captured by an exponential with offset, $r_k = b_{\text{offset}} \cdot m_{\text{offset}}^k + c_{\text{offset}}$. If the residual of this exponential with offset R_{offset}^2 was smaller than the residual of the MR model R_{exp}^2 by a factor of two, $H_{\text{offset}} = (2 \cdot R_{\text{offset}}^2 < R_{\text{exp}}^2)$, then the data set was rejected. The factor two punishes for the differences in degree of freedom: The residuals of a model with two free parameters (exponential with offset) instead of one (exponential only) can only be smaller.

Second, ramping of the drive can lead to overestimation of autocorrelation times (Fig. S1e). The comparison of the two fits introduced above serves as a consistency check, which was able to identify ramping: if the data are captured by a BP, both models should infer identical \hat{m} . Thus, a difference between \hat{m}_{exp} and \hat{m}_{offset} hints at the invalidity of MR estimation. Instead of \hat{m} , we compared the autocorrelation times $\hat{\tau}_{\text{offset}} = -1/\log \hat{m}_{\text{offset}}$ and $\hat{\tau}_{\text{exp}}$ obtained from both models, as the logarithmic scaling increases the sensitivity. If their relative difference was too large, then the data are inconsistent with a BP and MR estimation invalid: $H_{\tau} = (|\tau_{\text{exp}} - \tau_{\text{offset}}| / \min\{\tau_{\text{exp}}, \tau_{\text{offset}}\} > 2)$.

Third, when a system changes between different states of activity, e.g. up and down states, the drive rate $\langle h_t \rangle$ may experience sudden jumps. These can lead to spurious autocorrelation (Fig. S1f). To identify these trends resulting from non-stationary input h_t or from choosing too short data sets, we tested whether the sequence of r_k was fit better by a linear regression $r_k = q_1 k + q_2$ on the pairs (k, r_k) , than by the exponential relation (S6). If the residuals R_{lin}^2 were smaller than R_{exp}^2 : $H_{\text{lin}} = (R_{\text{lin}}^2 < R_{\text{exp}}^2)$, data were rejected.

Apart from non-stationarities, even Poisson activity ($m = 0$, $A_t = h_t$) with stationary rate may lead to a spurious overestimation of \hat{m} as well: for *subsampled* branching processes of *finite* duration, the Poisson case and processes close to criticality ($m = 1$) can show very similar autocorrelation results, because the sequence of r_k is expected to be absolutely or almost flat, respectively. Moreover, for $m = 0$ any solution on the manifold with $b = 0$ minimizes the residuals in Eq. (S6). Hence, the estimator for \hat{m} may yield any value depending on the initial conditions of the minimization scheme. To distinguish between $m = 0$ and $m > 0$, we used the fact that for $m = 0$, all slopes r_k are expected to be distributed around zero, $\langle r_k \rangle = 0$. In contrast, for processes with $m > 0$, all slopes are expected to be larger than zero $\langle r_k \rangle = b \cdot m^k > 0$. Thus to identify stationary Poisson activity, we tested (using a one-sided t-test) if the slopes obtained from the data were significantly larger than zero, yielding the p -value $p_{\bar{r} \leq 0}$ and the following test (Fig. S1b): $H_{\bar{r} \leq 0} = (p_{\bar{r} \leq 0} \geq 0.1)$. The choice of the significance level should be guided by the severity of type I or II errors here: if it is set too liberal, Poisson activity may be mistaken for correlated activity, potentially even close-to-critical. On the other hand, if the significance level is too conservative, activity with long autocorrelation times may be spuriously considered Poissonian under strong subsampling (when b is small and all slopes only slightly differ from zero). For this study, we chose a significance level of $p_{\bar{r} \leq 0} < 0.1$ in order to not underestimate the risk of large activity cascades. To confirm candidates for Poisson activity identified through positive $H_{\bar{r} \leq 0}$, we assured that the r_k did not show a systematic trend, i.e. that linear regression of r_k as a function of k (see H_{lin} above) yielded slope zero: $H_{q_1=0} = (p_{q_1=0} \geq 0.05)$. The according significance level for this two sided test is then given by $p_{q_1 \neq 0} < 0.05$.

We discriminate the following cases in the order indicated in Tab. S1: \hat{m} obtained from MR estimation is only valid if none of the tests (except $H_{q_1=0}$, which is ignored here) is positive. A positive result for any of H_{offset} , H_{τ} , or H_{lin} indicates non-stationarities, the data are not explained by a stationary BP, and MR estimation is invalid. If $H_{\bar{r} \leq 0}$ is positive, the data are potentially consistent with Poisson activity ($m = 0$). This is only the case if $H_{q_1=0}$ is also positive, if otherwise $H_{q_1=0}$ is negative, the Poisson hypothesis is also rejected and MR estimation invalid. This strategy correctly identified the validity of MR estimation for all investigated cases: stationary BPs with $m = 0.98$ and $m = 0.0$ were accepted, while nonstationary BPs with transient changes, ramping, or sudden jumps of the drive were excluded (Fig. S1).

Supp. 6 Variance of the estimates.

The distribution of \hat{m} is consistent with a normal distribution $\mathcal{N}(m, \sigma_m^2)$ centered around the true mean offspring m (Fig. S2a; numerical results). The variance σ_m^2 depends on the branching ratio m , the mean activity $\langle A_t \rangle$, the length L of the time series, and the sampling fraction α . Each of these factors affects σ_m^2 mainly by changing the *effective length* of the time series, i.e. the number of non-zero entries $l = |\{A_t | A_t > 0\}|$. Thus, regardless of the actual time series length L or the mean activity $\langle A_t \rangle$, the variance scales as a power-law in l , $\text{Var}[\hat{m}] \propto l^{-\gamma}$ (Fig. S2b). The exponent of this power-law depends on m . The closer to criticality the process is, the larger

H_{offset}	H_{τ}	H_{lin}	$H_{\bar{r} \leq 0}$	$(H_{q_1=0})$	interpretation	
\times	\times	\times	\times	$-$	BP with $m = \hat{m}$ explains data	MR estimation valid
\checkmark	$-$	$-$	$-$	$-$	data not explained by BP	MR estimation invalid
$-$	\checkmark	$-$	$-$	$-$		
$-$	$-$	\checkmark	$-$	$-$		
$-$	$-$	$-$	\checkmark	\times		
$-$	$-$	$-$	\checkmark	\checkmark	Poisson activity ($m = 0$) explains data	MR estimation valid

Table S1: Consistency checks for MR estimation. In order to assess if the results obtained from MR estimation are consistent with a BP with stationary parameters, we perform five tests (see supplementary text). We discriminate the following cases in this order: A BP with $m = \hat{m}$ is only considered to explain the data, if the four tests H_{offset} , H_{τ} , H_{lin} , and $H_{\bar{r} \leq 0}$ are negative (\times). If any of H_{offset} , H_{τ} , or H_{lin} is positive (\checkmark), the data cannot be explained by a BP with any m , regardless of the other tests ($-$), and MR estimation is invalid. If $H_{\bar{r} \leq 0}$ is positive, the additional test $H_{q_1=0}$ becomes relevant: if it is negative, the data cannot be explained by a BP with any m . If it is also positive, the data are consistent with Poisson activity (BP with $m = 0$).

the exponent γ , i.e. the larger benefit from longer time series length l . For $m = 0.99$, we found $\gamma \approx 3/2$. The performance of the estimator is in principle independent of the mean activity, small $\langle A_t \rangle$ only affect the variance of the MR estimator through a potential decrease of l .

Similarly, the degree of subsampling only affects the variance of the estimator through a decrease of the effective length of a_t . While there may be a significant rise in σ_m^2 when reducing the sampling fraction α , this increase can be explained by the coincidental decrease in l , as the rescaled variance $\sigma_m^2 \cdot l^\gamma$ remains within one order of magnitude over four decades of the sampling fraction α (Fig. S2c).

How does the variance change close to the critical transition? We found that the answer to this question highly depends on the specific choice of the parameters: if m is varied, one can either keep $\langle A_t \rangle$ or h constant, not both at the same time. If the mean activity $\langle A_t \rangle$ is fixed by choosing $h = \langle A_t \rangle (1 - m)$, then the variance of the process scales as $\text{Var}[A_t] \propto 1/(1 - m)$. As $m \rightarrow 1$, the activity will inevitably get into a regime, where bursts of activity ($A_t > 0$) are disrupted by intermittent quiescent periods ($A_t = 0$), thereby reducing l . In turn, the variance of the estimator increases as detailed before.

If however, the drive h is kept constant, we found that variance scales linearly in the distance to criticality $\epsilon = 1 - m$ over at least 5 orders of magnitude of ϵ : $\sigma_m^2 \propto \epsilon$ (Fig. S2d). Thus, the variance decreases when approaching criticality, while the relative variance σ_m^2/ϵ is constant. Note, however, that even though the standard deviation also decreases when approaching criticality ($\sigma_m \propto \sqrt{\epsilon}$), the relative standard deviation increases ($\sigma_m/\epsilon \propto 1/\sqrt{\epsilon}$). We obtained scaling laws with the same exponents for other quadratic (like the mean squared error MSE) and linear (like the inter-quartile range IQR) measures of variation.

Confidence interval estimation. We used a model based approach to estimate confidence intervals for both simulation and experimental data. For simulations, we simulated $B \in \mathbb{N}$ independent copies of the investigated model and applied MR estimation to each copy, yielding a collection of B independent estimates $\{\hat{m}^{(b)}\}_{b=1}^B$.

For experimental time series a_t with length L , mean activity $\langle a_t \rangle$, and number of sampled units n , MR estimation yields an estimate \hat{m} . We then simulated B copies of branching networks $\{A_t^{(b)}\}_{b=1}^B$ (for simulation details see Supp. 7) with $N = 10,000$ units, length L , $m = \hat{m}$ as inferred by MR estimation, and drive h . We hereby chose the drive $h = \langle a_t \rangle (1 - \hat{m}) N/n$ such that after subsampling n units, the mean activity of each resulting time series $a_t^{(b)}$ matched that of the original time series a_t , $\langle a_t^{(b)} \rangle = \langle a_t \rangle$. This procedure gives B copies of a BN that all match a_t in terms of the mean activity, the branching ratio, time series length, and number of sampled units. Applying MR estimation to these BNs yields a collection of B independent estimates $\{\hat{m}^{(b)}\}_{b=1}^B$. For both simulation and experimental data, the distribution of \hat{m} and confidence intervals can be constructed from this collection.

Supp. 7 Simulations

Branching process. We simulated BPs according to Eq. (S1) in the following way: Realizations of the random numbers $y_{t,i}$ and h_t describing the number of offsprings, and the drive, were each drawn from a Poisson distribution: $y_{t,i} \sim \text{Poi}(m)$ with mean m , and $h_t \sim \text{Poi}(h)$ with mean h , respectively. Here, we used Poisson distributions as they allow for non-trivial offspring distributions with easy control of the branching ratio m by only one parameter. For the brain, one might assume that each neuron is connected to k postsynaptic neurons, each of which is excited with probability p , motivating a binomial offspring distribution with mean $m = kp$. As in cortex k is typically large and p is typically small, the Poisson limit is a reasonable approximation. For the performance of the MR estimator and the limit behavior of the BP, the particular form of the law Y is not

important such that the special choice we made here does not restrict the generality of our results.

The mean rate $\langle A_t \rangle$ depends on m and h . In the simulation we varied m and fixed $\langle A_t \rangle = 100$ by adjusting h accordingly if not stated otherwise. For subsampling the BP, each unit is observed independently with probability $p \leq 1$. Then a_t is distributed following a binomial distribution $\text{Bin}(A_t, p)$, and subsampling is implemented by drawing a_t from A_t at each time step. As $\langle a_t \rangle = p A_t$, this implementation of subsampling satisfies the definition of stochastic subsampling with $\alpha = p$, $\beta = 0$.

Branching network. In addition to the classical branching process, we also simulated a branching network model (BN) by mapping a branching process [6, 21] onto a fully connected network of $N = 10000$ neurons. An active neuron activated each of its k postsynaptic neurons with probability $p = m/k$. Here, $k = 4$ postsynaptic neurons were drawn randomly without replacement at each step, thereby avoiding that two different active neurons would both activate the same target neuron. Similar to the BP, the BN is critical for $m = 1$ in the infinite size limit, and subcritical (supercritical) for $m < 1$ ($m > 1$). As detailed for the BP, h was adjusted to the choice of m to achieve $\langle A_t \rangle = 100$, which corresponds to a rate of 0.01 spikes per neuron and time step. Subsampling [4] was applied to the model by sampling the activity of n neurons only, which were selected randomly before the simulation, and neglecting the activity of all other neurons.

Self-organized critical model. The SOC neural network model we used here is the Bak-Tang-Wiesenfeld (BTW) model [26]. Translated to a neuroscience context, the model consisted of $N = 10000$ (100×100) non-leaky integrate and fire neurons. A neuron i spiked if its membrane voltage $V_i(t)$ reached a threshold θ

$$\text{If } V_i(t) > \theta, V_i(t+1) = V_i(t) - 4. \quad (\text{S16})$$

Note that the choice of θ does not change the activity of the model at all, so we set $\theta = 0$ for convenience. The model neurons were arranged on a 2D lattice, and each neuron was connected locally to its four nearest neighbors with coupling strength $\alpha_{ij} = \alpha$:

$$V_i(t+1) = V_i(t) + \sum_j \alpha_{ij} \delta(t - T_j) + h(t), \quad (\text{S17})$$

where T_j denotes the spike times of neuron j , and h_t is the Poisson drive with mean rate h as defined for the BP above. Note that the neurons at the edges and corners of the grid had only 3 and 2 neighbors, respectively. This model is equivalent to the well-known Bak-Tang-Wiesenfeld model [26] if $h \rightarrow 0$ and $\alpha = 1$. Subsampling [4] was applied to the model by sampling the activity of n neurons only, which were selected randomly before the simulation, and neglecting the activity of all other neurons.

Parameter choices. If not stated otherwise, simulations were run for $L = 10^7$ time steps or until A_t exceeded 10^9 , i.e. approximately half of the 32 bit integer range. If not stated otherwise, confidence intervals (see Supp. 6) were estimated from $B = 100$ samples, both for simulation and experiments.

In Figs. 1c,d, BNs and the BTW model were simulated with $N = 10^4$ units and $\langle A_t \rangle = 100$. In Fig. 1e, a BP was simulated with $m = 0.9$ and $\langle A_t \rangle = 100$.

In Fig. 3d, subcritical and critical BNs with $N = 10^4$ and $\langle A_t \rangle = 100$ were simulated, and $n = 100$ units sampled. Because of the non-stationary, exponential growth in the supercritical case, here BPs were simulated with $h = 0.1$ and units observed with probability $\alpha = 0.01$.

The *in vivo*-like BN was defined to match the cat recording, with $m = 0.98$, $n = 50$, and $\langle a_t \rangle = 1.58$ per bin, from which $h/N = 0.032$ per bin and neuron follows. Here, we chose $N = 10^4$. To test for stationarity, the cat recordings and the *in vivo*-like BN were split into 59 windows of 1250 time steps each, before estimating m for each window.

Supp. 8 Epidemiological recordings

WHO data on measles worldwide. Time series with yearly case reports for measles in 194 different countries are available online from the World Health Organization (WHO) for the years between 1980 and 2014. MR estimation was applied to these time series. Because they contain very few data points and potential long-term drifts, we applied the consistency checks detailed above for every country (Tab. S1). After these checks, 124 out of the 194 surveyed countries were accepted for MR analysis and included in our analysis. Yearly information on approximate vaccination percentages (measles containing vaccine dose 1, MCV1) for the same countries and time span are also available online from the WHO.

RKI data on norovirus, measles and MRSA in Germany. For Germany, the Robert-Koch-Institute (RKI) surveys a range of infectious diseases on a weekly basis, including measles, norovirus, and invasive meticillin-resistant *Staphylococcus aureus* (MRSA). Case reports are available through their SURVSTAT@RKI server [58]. Because of possible changes in report policies in the beginning of surveillance, we omitted the data from the first 6 months of each recording. Moreover, we omitted the incomplete week on the turn of the year, thus evaluating 52 full weeks in each year.

The MRSA recording showed a slow, small variation in the case reports that can be attributed to slow changes in the drive rates. To compensate for these slow drifts, we corrected the time series by subtracting a moving average over 3 years (156 weeks). We then applied MR estimation to the obtained time series. The recordings for measles and norovirus showed strong seasonal fluctuations of the case reports, resulting in a baseline oscillation of the autocorrelation function. We therefore used a modified model

$$r_k = b \cdot m^k + c_1 \cdot \cos(2\pi k/T) + c_2 \quad (\text{S18})$$

with a fixed period of $T = 52$ weeks, and estimated \hat{m} , \hat{b} , \hat{c}_1 , and \hat{c}_2 from minimizing the residual of this modified equation.

Supp. 9 Animal experiments

We evaluated spike population dynamics from recordings in rats, cats and monkeys. The rat experimental protocols were approved by the Institutional Animal Care and Use Committee of Rutgers University [51, 50]. The cat experiments were performed in accordance with guidelines established by the Canadian Council for Animal Care [49]. The monkey experiments were performed according to the German Law for the Protection of Experimental Animals, and were approved by the Regierungspräsidium Darmstadt. The procedures also conformed to the regulations issued by the NIH and the Society for Neuroscience. The spike recordings from the rats and the cats were obtained from the NSF-funded CRCNS data sharing website [59, 49, 51, 50].

In rats the spikes were recorded in CA1 of the right dorsal hippocampus during an open field task. We used the first two data sets of each recording group (ec013.527, ec013.528, ec014.277, ec014.333, ec015.041, ec015.047, ec016.397, ec016.430). The data-sets provided sorted spikes from 4 shanks (ec013) or 8 shanks (ec014, ec015, ec016), with 31 (ec013), 64 (ec014, ec015) or 55 (ec016) channels. We used both, spikes of single and multi units, because knowledge about the identity and the precise number of neurons is not required for the MR estimator. More details on the experimental procedure and the data-sets proper can be found in [51, 50].

For the spikes from the cat, neural data were recorded by Tim Blanche in the laboratory of Nicholas Swindale, University of British Columbia [49]. We used the data set pvc3, i.e. recordings in area 18 which contain 50 sorted single units [59]. We used that part of the experiment in which no stimuli were presented, i.e., the spikes reflected spontaneous activity in the visual cortex of the anesthetized cat. Because of potential non-stationarities at the beginning and end of the recording, we omitted data before 25 s and after 320 s of recording. Details on the experimental procedures and the data proper can be found in [49, 59].

The monkey data are the same as in [48, 28]. In these experiments, spikes were recorded simultaneously from up to 16 single-ended micro-electrodes ($\varnothing = 80 \mu\text{m}$) or tetrodes ($\varnothing = 96 \mu\text{m}$) in lateral prefrontal cortex of three trained macaque monkeys (M1: 6 kg ♀; M2: 12 kg ♂; M3: 8 kg ♀). The electrodes had impedances between 0.2 and 1.2 M Ω at 1 kHz, and were arranged in a square grid with inter electrode distances of either 0.5 or 1.0 mm. The monkeys performed a visual short term memory task. The task and the experimental procedure is detailed in [48]. We analyzed spike data from 11 experimental sessions comprising almost 12.000 trials (M1: 4 sessions; M2: 4 sessions; M3: 3 sessions). 6 out of 11 sessions were recorded with tetrodes. Spike sorting on the tetrode data was performed using a Bayesian optimal template matching approach as described in [60] using the Spyke Viewer software [61]. On the single electrode data, spikes were sorted with a multi-dimensional PCA method (Smart Spike Sorter by Nan-Hui Chen).

Analysis. For each recording, we collapsed the spike times of all recorded neurons into one single train of population spike counts a_t , where a_t denotes how many neurons spiked in the t^{th} time bin Δt . We used $\Delta t = 4$ ms, reflecting the propagation time of spikes from one neuron to the next.

From these time series, we estimated \hat{m} using the MR estimator with $k_{\text{max}} = 2500$ (corresponding to 10 s) for the rat recordings, $k_{\text{max}} = 150$ (600 ms) for the cat recording, and $k_{\text{max}} = 500$ (2000 ms) for the monkey recordings, assuring that k_{max} was always in the order of multiple autocorrelation times. For the monkey experiments, we performed MR estimation on the pooled data of all ~ 700 trials of 5 s length each. Experiments were excluded if the tests according to Supp. 5 detected potential nonstationarities.

References

- [1] Laure Papoz, Beverley Balkau, and Joseph Lellouch. Case counting in epidemiology: Limitations of methods based on multiple data sources. *Int. J. Epidemiol.*, 25(3):474–478, 1996.
- [2] RM Cormack. Problems with using capture-recapture in epidemiology: An Example of a Measles Epidemic. *J. Clin. Epidemiol.*, 52(10):909–14, 1999.
- [3] Mario Quagliariello. *Stress-testing the banking system: methodologies and applications*. Cambridge University Press, New York, 2009.
- [4] Viola Priesemann, Matthias H J Munk, and Michael Wibral. Subsampling effects in neuronal avalanche distributions recorded in vivo. *BMC Neurosci.*, 10:40, jan 2009.
- [5] Tiago L. Ribeiro, Mauro Copelli, Fábio Caixeta, Hindiael Belchior, Dante R. Chialvo, Miguel a L Nicolelis, and Sidarta Ribeiro. Spike Avalanches Exhibit Universal Dynamics across the Sleep-Wake Cycle. *PLoS One*, 5(11):e14129, nov 2010.
- [6] Theodore E. Harris. *The Theory of Branching Processes*. Springer Berlin, 1963.
- [7] C. R. Heathcote. A Branching Process Allowing Immigration. *J. R. Stat. Soc. Ser. B*, 27(1):138–143, 1965.
- [8] A. G. Pakes. Branching Processes with Immigration. *J. Appl. Probab.*, 8(1):32, mar 1971.
- [9] Harry Kesten. Random difference equations and Renewal theory for products of random matrices. *Acta Math.*, 131(1):207–248, dec 1973.
- [10] Håkan Andersson and Tom Britton. *Stochastic Epidemic Models and Their Statistical Analysis*, volume 151 of *Lecture Notes in Statistics*. Springer New York, New York, NY, 2000.
- [11] C P Farrington, M N Kanaan, and N J Gay. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, 4(2):279–95, 2003.
- [12] Markus Loeffler and Bernd Grossmann. A stochastic branching model with formation of subunits applied to the growth of intestinal crypts. *J. Theor. Biol.*, 150(2):175–191, may 1991.
- [13] Miguel González, Rodrigo Martínez, and Manuel Mota. Bisexual branching processes to model extinction conditions for Y-linked genes. *J. Theor. Biol.*, 258(3):478–88, jun 2009.
- [14] Marek Kimmel and David E. Axelrod. *Branching Processes in Biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY, 2015.
- [15] Amnon Pazy and PaulH. Rabinowitz. On a branching process in neutron transport theory. *Arch. Ration. Mech. Anal.*, 51(2):153–164, 1973.
- [16] Vladimir Filimonov and Didier Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Phys. Rev. E*, 85(5):056108, may 2012.
- [17] Fabio Caccioli, Munik Shrestha, Cristopher Moore, and J. Doynne Farmer. Stability analysis of financial contagion due to overlapping portfolios. *J. Bank. Financ.*, 46:233–245, sep 2014.
- [18] T.W. Epps. Stock prices as branching processes. *Commun. Stat. Stoch. Model.*, 12(4):529–558, jan 1996.
- [19] Georgi K. Mitov, S. T. Rachev, Y. S. Kim, and F. J. Fabozzi. Barrier Option Pricing by Branching Processes. *Int. J. Theor. Appl. Financ.*, 12(07):1055–1073, nov 2009.
- [20] John M Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. *J. Neurosci.*, 23(35):11167–11177, 2003.
- [21] Clayton Haldeman and John Beggs. Critical Branching Captures Activity in Living Neural Networks and Maximizes the Number of Metastable States. *Phys. Rev. Lett.*, 94(5):058101, feb 2005.
- [22] Hermann Cuntz, Friedrich Forstner, Alexander Borst, and Michael Häusser. One Rule to Grow Them All: A General Theory of Neuronal Branching and Its Practical Application. *PLoS Comput. Biol.*, 6(8):e1000877, 2010.
- [23] C. C. Heyde and E. Seneta. Estimation Theory for Growth and Immigration Rates in a Multiplicative Process. *J. Appl. Probab.*, 9(2):235, jun 1972.
- [24] C.Z. Wei and J. Winnicki. Estimation of the Means in the Branching Process with Immigration. *Ann. Stat.*, 18(4):1757–1773, dec 1990.
- [25] Adiel Statman, Maya Kaufman, Amir Minerbi, Noam E Ziv, and Naama Brenner. Synaptic Size Dynamics as an Effectively Stochastic Process. *PLoS Comput. Biol.*, 10(10):e1003846, oct 2014.
- [26] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.*, 59(4):381–384, jul 1987.
- [27] Viola Priesemann, Mario Valderrama, Michael Wibral, and Michel Le Van Quyen. Neuronal avalanches differ from wakefulness to deep sleep—evidence from intracranial depth recordings in humans. *PLoS Comput. Biol.*, 9(3):e1002985, jan 2013.
- [28] Viola Priesemann, Michael Wibral, Mario Valderrama, Robert Pröpper, Michel Le Van Quyen, Theo Geisel, Jochen Triesch, Danko Nikolić, and Matthias H J Munk. Spike avalanches in vivo suggest a driven, slightly subcritical brain state. *Front. Syst. Neurosci.*, 8(June):108, jan 2014.
- [29] Anja M. Hauri, Hans-Jürgen Westbrock, Herman Claus, Steffen Geis, Siegfried Giernat, Michael Forßbohm, and Helmut Uphoff. Electronic Outbreak Surveillance in Germany: A First Evaluation for Nosocomial Norovirus Outbreaks. *PLoS One*, 6(3):e17341, mar 2011.
- [30] Helen W. Boucher and G. Ralph Corey. Epidemiology of MethicillinResistant Staphylococcus aureus. *Clin. Infect. Dis.*, 46(S5):S344–S349, 2008.

- [31] Madeleine Dulon, Frank Haamann, Claudia Peters, Anja Schablon, and Albert Nienhaus. MRSA prevalence in european healthcare settings: a review. *BMC Infect. Dis.*, 11(1):138, 2011.
- [32] Peter F.M. Teunis, Christine L. Moe, Pengbo Liu, Sara E. Miller, Lisa Lindesmith, Ralph S. Baric, Jacques Le Pendu, and Rebecca L. Calderon. Norwalk virus: How infectious is it? *J. Med. Virol.*, 80(8):1468–1476, aug 2008.
- [33] Robin Köck, Alexander Mellmann, Frieder Schaumburg, Alexander W Friedrich, Frank Kipp, and Karsten Becker. The epidemiology of methicillin-resistant *Staphylococcus aureus* (MRSA) in Germany. *Dtsch. Arztebl. Int.*, 108(45):761–7, 2011.
- [34] Frank R. DeLeo, Michael Otto, Barry N. Kreiswirth, and Henry F. Chambers. Community-associated methicillin-resistant *Staphylococcus aureus*. *Lancet*, 375(9725):1557–1568, 2010.
- [35] C. v. Vreeswijk and H Sompolinsky. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science (80-.)*, 274(5293):1724–1726, dec 1996.
- [36] Nicolas Brunel. Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *J. Physiol. Paris*, 94(5-6):445–463, 2000.
- [37] Alfonso Renart, J. de la Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D Harris. The Asynchronous State in Cortical Circuits. *Science (80-.)*, 327(5965):587–590, jan 2010.
- [38] a. Levina, J. M. Herrmann, and T. Geisel. Dynamical synapses causing self-organized criticality in neural networks. *Nat. Phys.*, 3(12):857–860, nov 2007.
- [39] Daniel Millman, Stefan Mihalas, Alfredo Kirkwood, and Ernst Niebur. Self-organized criticality occurs in non-conservative neuronal networks during up’ states. *Nat. Phys.*, 6(10):801–805, 2010.
- [40] Thierry Mora and William Bialek. Are Biological Systems Poised at Criticality? *J. Stat. Phys.*, 144(2):268–302, 2011.
- [41] David J Schwab, Ilya Nemenman, and Pankaj Mehta. Zipf’s Law and Criticality in Multivariate Data without Fine-Tuning. *Phys. Rev. Lett.*, 113(6):068102—, aug 2014.
- [42] Nils Bertschinger and Thomas Natschläger. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput.*, 16(7):1413–1436, 2004.
- [43] Joschka Boedecker, Oliver Obst, Joseph T Lizier, N Michael Mayer, and Minoru Asada. Information processing in echo state networks at the edge of chaos. *Theory Biosci.*, 131(3):205–213, 2012.
- [44] Woodrow L Shew and Dietmar Plenz. The functional benefits of criticality in the cortex. *Neuroscientist*, 19(1):88–100, feb 2013.
- [45] Horace B Barlow. Possible Principles Underlying the Transformations of Sensory Messages. In Walter A. Rosenblith, editor, *Sens. Commun.*, pages 217—234. The MIT Press, sep 2012.
- [46] Joseph J Atick. Could information theory provide an ecological theory of sensory processing? *Netw. Comput. neural Syst.*, 3(2):213–251, 1992.
- [47] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2000.
- [48] Gordon Pipa, Ellen S Städtler, Eugenio F Rodriguez, James A Waltz, Lars F Muckli, Wolf Singer, Rainer Goebel, and Matthias H J Munk. Performance- and stimulus-dependent oscillations in monkey prefrontal cortex during short-term memory. *Front. Integr. Neurosci.*, 3(October):25, jan 2009.
- [49] Tim Blanche. Multi-neuron recordings in primary visual cortex, 2009.
- [50] K Mizuseki, A Sirota, E Pastalkova, and G. Buzsáki. Theta Oscillations Provide Temporal Windows for Local Circuit Computation in the Entorhinal-Hippocampal Loop. *Neuron*, 64(2):267–280, 2009.
- [51] K Mizuseki, A Sirota, E Pastalkova, and G. Buzsáki. Multi-unit recordings from the rat hippocampus made during open field foraging, 2009.
- [52] John D Murray, Alberto Bernacchia, David J Freedman, Ranulfo Romo, Jonathan D Wallis, Xinying Cai, Camillo Padoa-Schioppa, Tatiana Pasternak, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.*, 17(12):1661–3, 2014.
- [53] Márton Ispány and Gyula Pap. Critical branching processes with immigration. In Miguel González Velasco, Inés M. Puerto, Rodrigo Martínez, Manuel Molina, Manuel Mota, and Alfonso Ramos, editors, *Lect. Notes Stat.*, volume 197, pages 135–146. Springer Berlin Heidelberg, 2010.
- [54] A. A. Alzaid and M. Al-Osh. An Integer-Valued pth-Order Autoregressive Structure (INAR(p)) Process. *J. Appl. Probab.*, 27(2):314, jun 1990.
- [55] M. Kachour and J. F. Yao. First-order rounded integer-valued autoregressive (RINAR(1)) process. *J. Time Ser. Anal.*, 30(4):417–448, 2009.
- [56] K. N. Venkataraman. A Time Series Approach to the Study of the Simple Subcritical Galton-Watson Process with Immigration. *Adv. Appl. Probab.*, 14(1):1–20, 1982.
- [57] J F Kenney and E S Keeping. Linear regression and correlation. *Math. Stat.*, 1:252–285, 1962.
- [58] Robert-Koch-Institute. SurvStat@RKI 2.0.
- [59] Timothy J Blanche and Nicholas V Swindale. Nyquist interpolation improves neuron yield in multiunit recordings. *J. Neurosci. Methods*, 155(1):81–91, jul 2006.
- [60] Felix Franke, Michal Natora, Clemens Boucsein, Matthias H J Munk, and Klaus Obermayer. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *J. Comput. Neurosci.*, 29(1-2):127–48, aug 2010.
- [61] Robert Pröpper and Klaus Obermayer. Spyke Viewer: a flexible and extensible platform for electrophysiological data analysis. *Front. Neuroinform.*, 7(November):26, jan 2013.

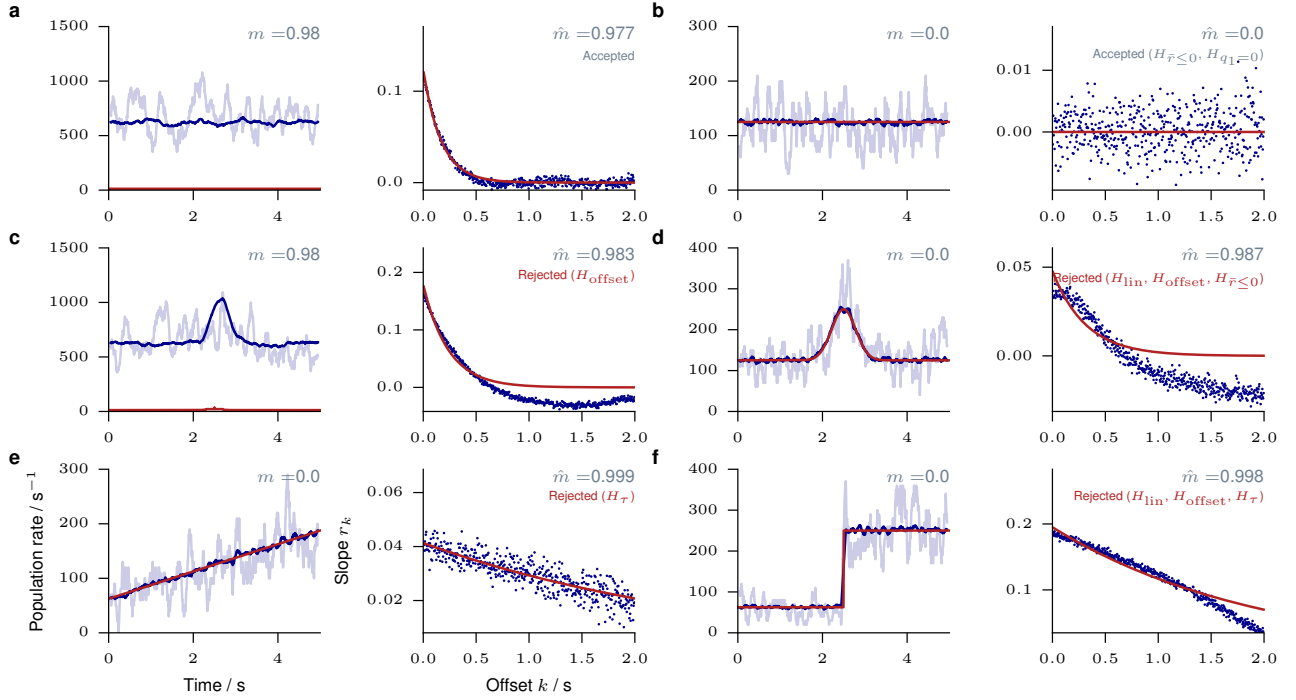


Figure S1: Excluding nonstationary data. Each left panels shows the time series a_t of the activity from one single trial (light blue) and averaged activity from 100 trials (dark blue), recorded from $n = 50$ out of $N = 10^4$ neurons. Each right panels shows the corresponding MR estimation from one single trial. We investigated the following, generic cases for the temporal evolution of the drive rate $\langle h_t \rangle$: **a, b.** The drive is stationary ($\langle h_t \rangle$ identical for all t , red), so are the mean rates $\langle a_t \rangle$. **c, d.** The drive exhibits a transient increase centered around half of the simulation time. The mean rate $\langle a_t \rangle$ is therefore also time-dependent and follows the temporal evolution of $\langle h_t \rangle$. **e.** The drive shows a linear increase over the simulation. **f.** The drive exhibits a step function after half the simulation. Nonstationarities (**c – f**) typically lead to an overestimation of \hat{m} , which is particularly severe if the underlying dynamics is Poissonian ($m = 0$). The tests defined in Supp. 5 (see Tab. S1) were able to exclude time series where the investigated nonstationarities were present, while accepting the stationary cases **a, b**.

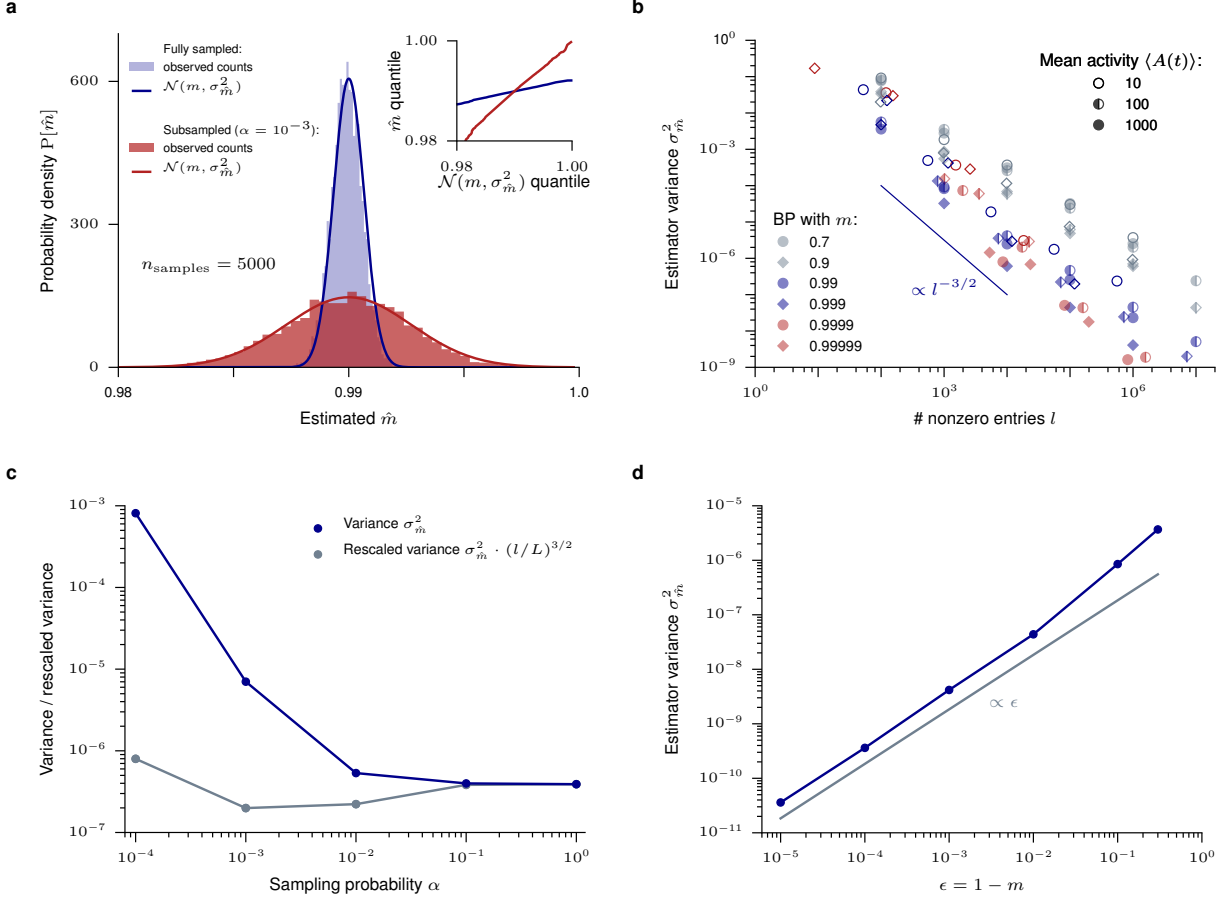


Figure S2: Variance of the MR estimates. This figure shows numerical result for the distribution and variability of the estimate \hat{m} as a function of multiple parameters. **a.** Distribution of the estimate \hat{m} , estimated from 5000 independent copies of a branching process (BP) with $m = 0.99$, $\langle A_t \rangle = 100$ and length $L = 10^5$: normalized histograms of the probability of estimating \hat{m} for full sampling (blue) and binomial subsampling with $\alpha = 0.001$ (red), together with normal distributions $\mathcal{N}(m, \sigma_m^2)$. Inset: Q-Q-plot for the quantiles of $\mathcal{N}(m, \sigma_m^2)$ and the quantiles of the estimated \hat{m} under both samplings. The estimated \hat{m} are found to be distributed normally in both cases (fully sampled: $r^2 = 0.9995$, subsampled: $r^2 = 0.998$). **b.** The variance $\sigma_{\hat{m}}^2$ of the estimate \hat{m} is estimated from 100 independent copies of a BP. Results for different m , mean activities $\langle A_t \rangle$ and time series lengths L are plotted as a function of the effective time series length $l = |\{A_t | A_t > 0\}|$, the number of nonzero entries. For any given m , the variance of \hat{m} shows algebraic scaling $\sigma_{\hat{m}}^2 \propto l^\gamma$. The exponent of this scaling depends on m , with higher γ the closer m is to unity. Hence, the benefit from longer time series is larger the closer a system is to criticality. Importantly, the variance does not directly depend on the mean activity $\langle A_t \rangle$, this number only influences the accuracy of MR estimation via the potential change in l . **c.** The variance of the estimate \hat{m} is estimated from 100 independent copies of a BP with $m = 0.99$, $\langle A_t \rangle = 100$, and $L = 10^5$ and plotted as a function of the sampling probability α under binomial subsampling. While the variance appears to increase dramatically under stronger subsampling, this increase can be attributed to the according decrease of the effective time series length l . After rescaling by $(l/L)^{3/2}$ (cf. panel **b**), the rescaled variance remains within one order of magnitude over four orders of magnitude in α . Hence, the accuracy of the estimator is not directly influenced by the degree of subsampling. **d.** The variance $\sigma_{\hat{m}}^2$ is estimated from 100 independent copies of a BP with $m = 0.99$, $h = 1$, and $L = 10^5$ and plotted as function of the distance to criticality $\epsilon = 1 - m$. The variance is found numerically to scale as $\sigma_{\hat{m}}^2 \propto \epsilon$, hence the standard deviation scales as $\sigma_{\hat{m}} \propto \sqrt{\epsilon}$. Similar scaling results were found for other linear (like the interquartile range) and quadratic (like the mean squared error) measures of variation.

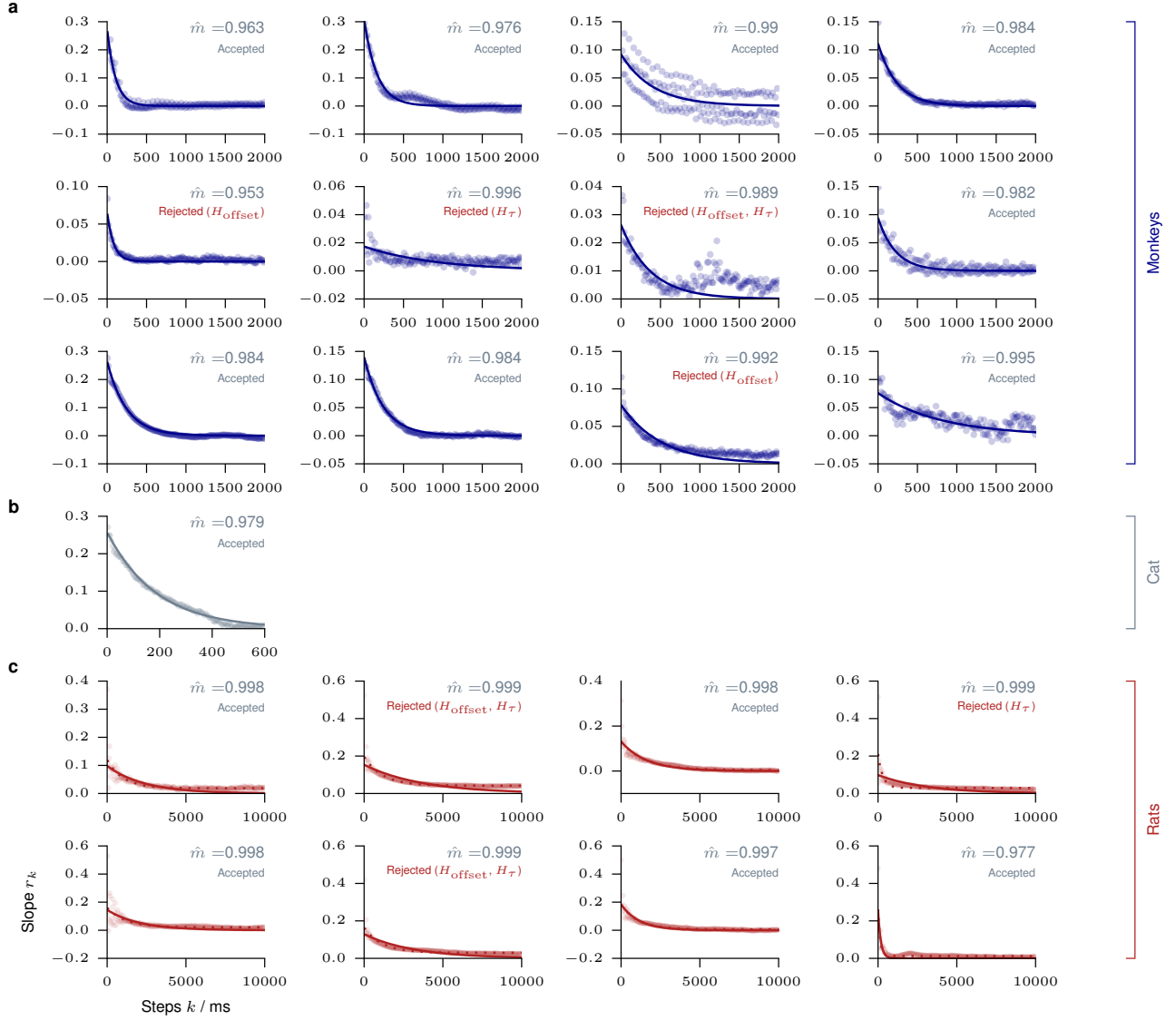


Figure S3: MR estimation for individual animals. MR estimation is shown for every individual animal (see Supp. 9). The consistency checks are detailed in the Supp. 5 (see Tab. S1). **a.** Data from monkey prefrontal cortex during an working memory task. The third panel shows a oscillation of r_k with a frequency of 50 Hz, corresponding to measurement corruption due to power supply frequency. **b.** Data from anesthetized cat primary visual cortex. **c.** Data from rat hippocampus during a foreaging task. In addition to a slow exponential decay, the slopes r_k show the ϑ -oscillations of 6 – 10 Hz present in hippocampus.